HMDD v2.0: a database for experimentally supported human microRNA and disease associations

Yang Li[1,2], Chengxiang Qiu[3], Jian Tu[2,4], Bin Geng[4], Jichun Yang[4], Tianzi Jiang[3,*] and Qinghua Cui[2,4,*]

[1]Department of Cell Biology, School of Basic Medical Sciences, Peking University, 38 Xueyuan Road, Beijing 100191, China, [2]Department of Biomedical Informatics, School of Basic Medical Sciences, Peking University, 38 Xueyuan Road, Beijing 100191, China, [3]Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China and [4]MOE Key Lab of Cardiovascular Sciences, Peking University, 38 Xueyuan Road, Beijing 100191, China

ABSTRACT

The Human microRNA Disease Database (HMDD; available via the Web site at http://cmbi.bjmu.edu. cn/hmdd and http://202.38.126.151/hmdd/tools/ hmdd2.html) is a collection of experimentally supported human microRNA (miRNA) and disease associations. Here, we describe the HMDD v2.0 update that presented several novel options for users to facilitate exploration of the data in the database. In the updated database, miRNA–disease association data were annotated in more details. For example, miRNA–disease association data from genetics, epigenetics, circulating miRNAs and miRNA–target interactions were integrated into the database. In addition, HMDD v2.0 presented more data that were generated based on concepts derived from the miRNA–disease association data, including disease spectrum width of miRNAs and miRNA spectrum width of human diseases. Moreover, we provided users a link to download all the data in the HMDD v2.0 and a link to submit novel data into the database. Meanwhile, we also maintained the old version of HMDD. By keeping data sets up-to-date, HMDD should continue to serve as a valuable resource for investigating the roles of miRNAs in human disease.

INTRODUCTION

MicroRNAs (miRNAs) are one class of important small noncoding RNA molecules that mainly repress gene expression at the posttranscription level (1). Generally, one miRNA could regulate hundreds of target genes, and one gene could also be regulated by hundreds of miRNAs (2). So far, ~2000 miRNAs have been identified in the human genome according to the miRBase database release 20 (3). Recently, increasing studies have shown that miRNAs play critical roles in many important biological processes (4). Therefore, miRNA-related dysfunctions could be associated with a broad spectrum of diseases, including cancer (5) and cardiovascular diseases (6). Clearly, miRNAs have been becoming one novel class of potential biomarkers or targets for disease diagnosis and therapy (7).

To date, a number of miRNA-related databases have been developed. These databases have shown their great helps in providing valuable miRNA-related information such as sequences (3), experimentally supported miRNA targets (8–10), mutations (11–13), experimentally supported miRNA transcription factors (14,15), miRNA–drug interactions (16,17) and miRNA-associated diseases (18–21). For example, miR2Disease is a database for experimentally miRNA–disease associations and miRNA–target interactions. miRTarBase is a database for experimentally validated miRNA–target interactions. Both databases have provided great help in miRNA-related studies.

To our knowledge, the Human microRNA Disease Database (HMDD), which was released in December 2007 and had been updated ~30 times during the past 5 years, is one of the first databases for miRNA-associated diseases (18). Here, we introduced the HMDD v2.0 that collected >10 000 experimentally supported miRNA–disease association entries, including ~600 miRNA genes and ~400 human diseases from >3000 articles. We also

annotated specific classes of entries, for example, entries whose experimental evidence is from genetics, epigenetics, circulating miRNAs and miRNA–target interactions. In addition, an analysis for the tendency of miRNA–disease investigations was performed. Finally, we summarized the usage of data sets in HMDD v2.0.

SYSTEM OVERVIEW

The aim of HMDD v2.0 is to provide a web interface for users to browse, search and download data sets in the database, and submit novel data into the database. To collect the experimentally supported miRNA–disease associations, we firstly obtained all miRNA-related publications from the PubMed database using the keywords 'microRNA', 'miRNA' or 'miR'. Then, we manually retrieved entries related with miRNA–disease associations. Every entry contains four items, which are miRNA name, disease name, experimental evidence for the miRNA–disease association and the publication PubMed ID. The miRNA name and the disease name are normalized. We further annotate the data in more details including entries whose experimental evidence is from genetics, epigenetics and miRNA–target interaction. More recently, increasing studies have revealed that a number of miRNAs stably exist in circulation systems and could be biomarkers for disease diagnosis and treatment (22). Therefore, we also annotated miRNA–disease entries whose evidence is from circulating miRNAs in HMDD v2.0. In addition, we integrated disease-related miRNA–target interactions from miR2Disease, miRTarBase and TarBase. As a result, HMDD collected 10 368 entries that include 572 miRNA genes, 378 diseases from 3511 articles.

In the 'HMDD v2.0' database, all data had been organized using SQLite, a lightweight database management system. The Web site was developed based on Django, a Python web framework. The database is available at http://cmbi.bjmu.edu.cn/hmdd and http://202.38.126.151/hmdd/tools/hmdd2.html.

QUERYING THE DATABASE

We provided users several ways to query the HMDD v2.0 database. First, users can browse the HMDD v2.0 by miRNA names or disease names. When clicking one miRNA or disease in the 'Browse' page, HMDD v2.0 will return a list of matched entries. Second, we provided a 'fuzzy search' function for the entries by the full or partial names of miRNAs or diseases in the 'Search' page. The 'Search' is case-insensitive. Moreover, all data in the database can be freely downloaded. The users can also submit novel data into the database. In addition, a detailed tutorial for the usage of the database is available in the 'Help' page.

THE USAGE OF THE DATA SETS IN HMDD v2.0

Besides general database search, browse and download as introduced above, users can perform their specific researches based on the datasets in HMDD. Here, we introduced a new analysis of HMDD data sets and summarized previous HMDD-based researches.

History tendency of miRNA–disease relationship investigations

By analyzing the publication time distribution of the total entries, entries from genetics, entries from epigenetics, entries from circulating miRNAs and entries from miRNA–target interactions, respectively, we showed that the total publications about miRNA–disease relationships increase dramatically (Figure 1A). This also suggests that the investigation of miRNAs in the pathogenesis of human disease is continually to be one of the hottest fields in biomedical research. However, the above four types of specific entries show different patterns. Entries from genetics (Figure 1B) increased dramatically from 2002 to 2007, and then increased slowly after 2007. In contrast, entries from epigenetics (Figure 1C), circulating miRNAs (Figure 1D) and miRNA–target interactions (Figure 1E) increased dramatically in recent years, suggesting that they are hot topics in the current miRNA–disease study. Especially for circulating miRNAs, there is no entry for circulating miRNA and disease relationship before 2009, but their entries increased dramatically after 2009, suggesting that establishment of circulating miRNAs as biomarkers for diagnosis and treatment of diseases has been becoming a hot topic in miRNA research.

Predicting miRNA-associated functions, disease and environmental factors

The miRNA–disease association data in HMDD can also be used to predict novel miRNA-associated diseases and functions. The hypothesis is that miRNAs with similar functions tend to be associated with diseases with similar phenotypes, and vice versa (18). Using HMDD miRNA–disease association data, we had previously developed a graph-based method to evaluate the functional similarity of miRNAs and then to infer novel miRNA-associated disease (23). Based on the miRNA functional similarity, two other labs developed network-based methods to predict novel miRNA-associated diseases (24,25). In addition, it is also possible to predict the relationship between environmental factors and diseases by calculating the similarity of their miRNA signatures in HMDD v2.0 (26).

miRNA set enrichment analysis

A miRNA set is defined as a group of miRNAs that have some specific features (18). According to this rule, miRNAs that are associated with the same disease, for example, breast cancer, will be presented as one miRNA set. As a result, HMDD can generate ∼400 miRNA sets. Using the miRNA enrichment analysis tool, TAM (27), it is easy to investigate enriched miRNA sets for a given list of miRNAs, for example, the upregulated miRNAs from a microarray experiment. This analysis makes it possible not only for finding patterns or rules behind a set of miRNAs but also for predicting novel miRNA–disease associations.
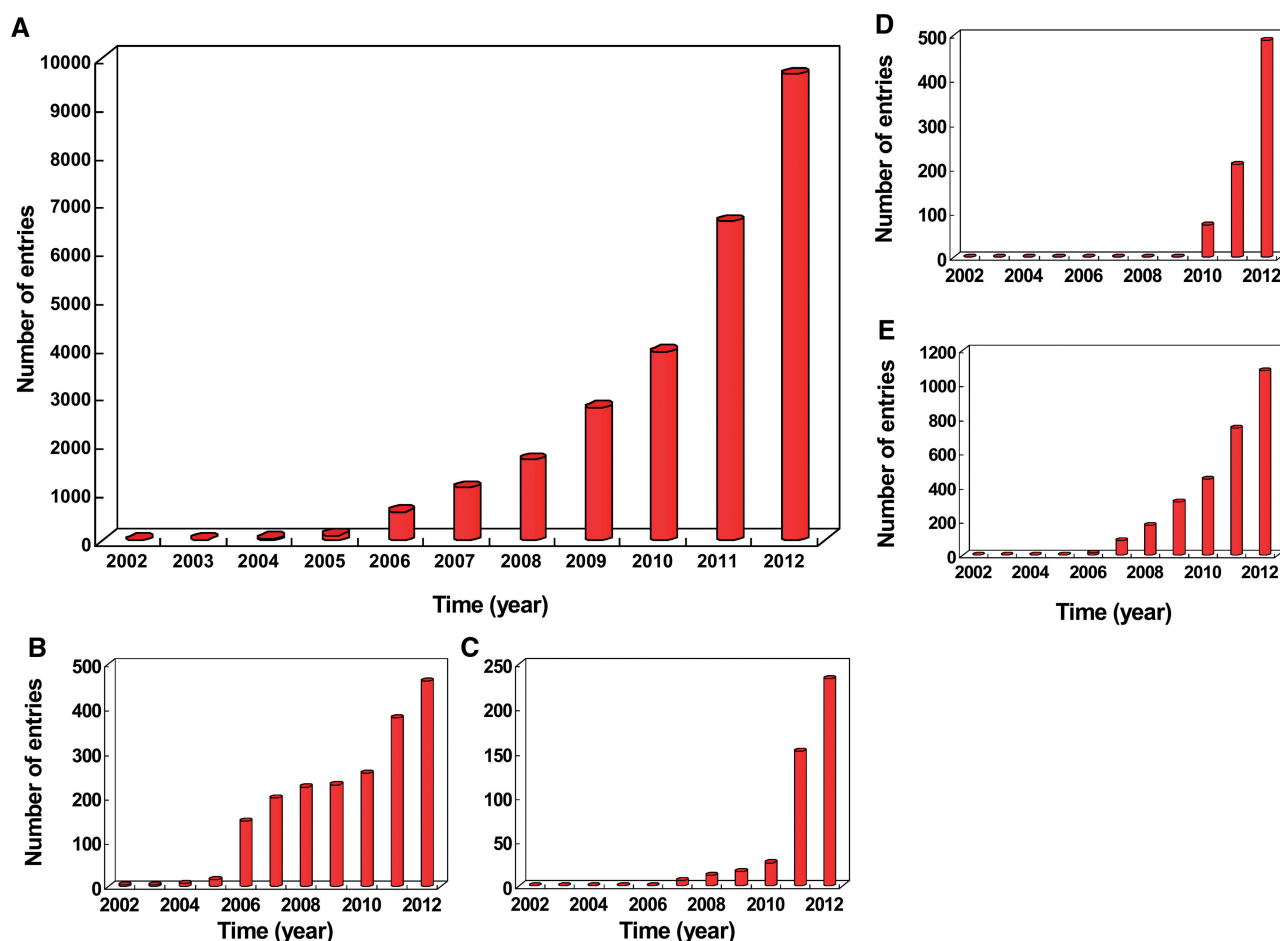
Figure 1. The historically cumulative number of total entries (**A**), entries from genetics (**B**), entries from epigenetics (**C**), entries from circulating miRNAs (**D**) and entries from miRNA–target interaction (**E**).

Disease spectrum width of a miRNA and miRNA spectrum width of a disease

The concept of disease spectrum width (DSW) of a miRNA was originally proposed by us in a previous study (26). For one miRNA i, $DSW(i) = n(i)/N$ where, $n(i)$ is the number of diseases associated with miRNA i, N is the total number of diseases that have been reported to be associated with miRNAs (26). We had previously shown that DSW of one miRNA could be a metric to evaluate its importance in function and human disease (26). For example, miR-21 has the biggest DSW (0.33), suggesting that it has a wide disease spectrum and plays an important role in many diseases. The top 10 miRNAs with the biggest DSWs are listed in Figure 2A. All of these miRNAs had been widely accepted to have critical functions and roles in human diseases. Using the similar procedure as described above, we also introduced another novel concept and metric for one disease, the miRNA spectrum width (MSW) of a disease. This metric could be used to evaluate the severity of a given disease. For example, all of the top 10 diseases (Figure 2B) with the biggest MSWs are among the most lethal human diseases. In future, more research cases are needed to confirm and consolidate the usefulness of the two

metrics, and we provided downloadable files for them in HMDD v2.0.

CONCLUSION

Increasing studies have shown that miRNAs have important functions and are associated with the development and progression of a broad range of diseases. miRNAs had been becoming novel potential molecules for disease diagnosis, treatment and prognosis. In this article, we describe an update (HMDD v2.0) of the HMDD. The HMDD database integrated experimentally supported miRNA–disease association data and further annotated four types of miRNA–disease association data. Based on miRNA–disease associations, we proposed and calculated two miRNA-related metrics, DSW of miRNAs and MSW of diseases, and provided downloadable files for them. By analyzing the publication time regarding the miRNA–disease associations in the past decade included in HMDD, we showed that this simple analysis could predict the tendency of miRNA–disease relationship investigations. For example, entries from circulating miRNAs increase most dramatically in recent years. According to this result, it is expected that more data
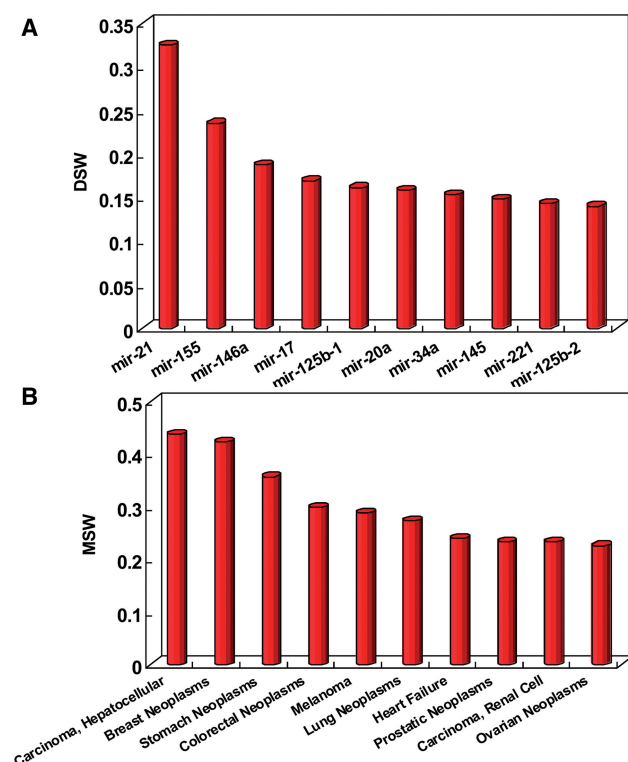
Figure 2. The top 10 miRNAs with the biggest DSW (**A**) and the top 10 diseases with the biggest MSW (**B**).

regarding the associations of circulating miRNAs with human diseases will be generated in the coming years. This suggests that the identification of circulating miRNAs as disease biomarkers is one of the hottest topics in miRNA research. In addition, we also provided evidence that HMDD data sets are also useful for predicting miRNA-associated diseases, functions and environmental factor–disease relationships. The important roles of miRNAs in diseases are attracting more biomedical researchers. Therefore, it is expected that HMDD v2.0 will integrate more experimentally supported miRNA–disease associations in the future. Finally, we believe that HMDD v2.0 is useful for the studies of miRNA–disease associations, and will provide more help in this field when it integrates more data and tools in the future.

REFERENCES

1. Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
2. Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
3. Griffiths-Jones,S. (2004) The microRNA Registry. *Nucleic Acids Res.*, **32**, D109–D111.
4. Ambros,V. (2004) The functions of animal microRNAs. *Nature*, **431**, 350–355.
5. Esquela-Kerscher,A. and Slack,F.J. (2006) Oncomirs - microRNAs with a role in cancer. *Nat. Rev. Cancer*, **6**, 259–269.
6. Latronico,M.V., Catalucci,D. and Condorelli,G. (2007) Emerging role of microRNAs in cardiovascular biology. *Circ. Res.*, **101**, 1225–1236.
7. Calin,G.A. and Croce,C.M. (2006) MicroRNA signatures in human cancers. *Nat. Rev. Cancer*, **6**, 857–866.
8. Hsu,S.D., Lin,F.M., Wu,W.Y., Liang,C., Huang,W.C., Chan,W.L., Tsai,W.T., Chen,G.Z., Lee,C.J., Chiu,C.M. *et al.* (2011) miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res.*, **39**, D163–D169.
9. Vergoulis,T., Vlachos,I.S., Alexiou,P., Georgakilas,G., Maragkakis,M., Reczko,M., Gerangelos,S., Koziris,N., Dalamagas,T. and Hatzigeorgiou,A.G. (2012) TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. *Nucleic Acids Res.*, **40**, D222–D229.
10. Xiao,F., Zuo,Z., Cai,G., Kang,S., Gao,X. and Li,T. (2009) miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res.*, **37**, D105–D110.
11. Gong,J., Tong,Y., Zhang,H.M., Wang,K., Hu,T., Shan,G., Sun,J. and Guo,A.Y. (2012) Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. *Hum. Mutat.*, **33**, 254–263.
12. Bhattacharya,A., Ziebarth,J.D. and Cui,Y. (2013) SomamiR: a database for somatic mutations impacting microRNA function in cancer. *Nucleic Acids Res.*, **41**, D977–D982.
13. Ziebarth,J.D., Bhattacharya,A., Chen,A. and Cui,Y. (2012) PolymiRTS Database 2.0: linking polymorphisms in microRNA target sites with human diseases and complex traits. *Nucleic Acids Res.*, **40**, D216–D221.
14. Wang,J., Lu,M., Qiu,C. and Cui,Q. (2010) TransmiR: a transcription factor-microRNA regulation database. *Nucleic Acids Res.*, **38**, D119–D122.
15. Yang,J.H., Li,J.H., Jiang,S., Zhou,H. and Qu,L.H. (2013) ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. *Nucleic Acids Res.*, **41**, D177–D187.
16. Liu,X., Wang,S., Meng,F., Wang,J., Zhang,Y., Dai,E., Yu,X., Li,X. and Jiang,W. (2012) SM2miR: a database of the experimentally validated small molecules' effects on microRNA expression. *Bioinformatics*, **29**, 409–411.
17. Yang,Q., Qiu,C., Yang,J., Wu,Q. and Cui,Q. (2011) miREnvironment database: providing a bridge for microRNAs, environmental factors and phenotypes. *Bioinformatics*, **27**, 3329–3330.
18. Lu,M., Zhang,Q., Deng,M., Miao,J., Guo,Y., Gao,W. and Cui,Q. (2008) An analysis of human microRNA and disease associations. *PLoS One*, **3**, e3420.
19. Yang,Z., Ren,F., Liu,C., He,S., Sun,G., Gao,Q., Yao,L., Zhang,Y., Miao,R., Cao,Y. *et al.* (2010) dbDEMC: a database of differentially expressed miRNAs in human cancers. *BMC Genomics*, **11(Suppl. 4)**, S5.
20. Jiang,Q., Wang,Y., Hao,Y., Juan,L., Teng,M., Zhang,X., Li,M., Wang,G. and Liu,Y. (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.*, **37**, D98–D104.
21. Ruepp,A., Kowarsch,A., Schmidl,D., Buggenthin,F., Brauner,B., Dunger,I., Fobo,G., Frishman,G., Montrone,C. and Theis,F.J. (2010) PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes. *Genome Biol.*, **11**, R6.
22. Creemers,E.E., Tijsen,A.J. and Pinto,Y.M. (2012) Circulating microRNAs: novel biomarkers and extracellular communicators in cardiovascular disease? *Circ Res.*, **110**, 483–495.
23. Wang,D., Wang,J., Lu,M., Song,F. and Cui,Q. (2010) Inferring the human microRNA functional similarity and functional

network based on microRNA-associated diseases. *Bioinformatics*, **26**, 1644–1650.

24. Chen,H. and Zhang,Z. (2013) Similarity-based methods for potential human microRNA-disease association prediction. *BMC Med. Genomics*, **6**, 12.

25. Chen,X., Liu,M.X. and Yan,G.Y. (2012) RWRMDA: predicting novel human microRNA-disease associations. *Mol. Biosyst.*, **8**, 2792–2798.

26. Qiu,C., Chen,G. and Cui,Q. (2012) Towards the understanding of microRNA and environmental factor interactions and their relationships to human diseases. *Sci. Rep.*, **2**, 318.

27. Lu,M., Shi,B., Wang,J., Cao,Q. and Cui,Q. (2010) TAM: a method for enrichment and depletion analysis of a microRNA category in a list of microRNAs. *BMC Bioinformatics*, **11**, 419.